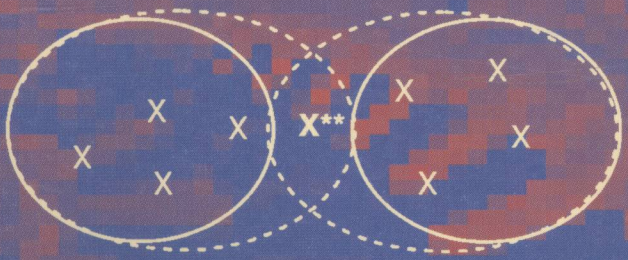
State-of-the-Art Survey

LNAI 3366

Iyad Rahwan
Pavlos Moraitis
Chris Reed (Eds.)

Argumentation in Multi-Agent Systems

**First International Workshop, ArgMAS 2004
New York, NY, USA, July 2004
Revised Selected and Invited Papers**



Springer

Iyad Rahwan Pavlos Moraitis
Chris Reed (Eds.)

Argumentation in Multi-Agent Systems

First International Workshop, ArgMAS 2004
New York, NY, USA, July 19, 2004
Revised Selected and Invited Papers

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Iyad Rahwan
The British University in Dubai, Institute of Informatics
P.O. Box 502216, Dubai, United Arab Emirates
E-mail: irahwan@acm.org
University of Melbourne, Dept. of Information Systems
E-mail: i.rahwan@pgrad.unimelb.edu.au

Pavlos Moraitis
University of Cyprus, Dept. of Computer Science
75 Kallipoleos Str., 1678 Nicosia, Cyprus
E-mail: moraitis@cs.ucy.ac.cy

Chris Reed
University of Dundee, Division of Applied Computing
Dundee DD1 4HN, Scotland, UK
E-mail: chris@computing.dundee.ac.uk

Library of Congress Control Number: 2004118426

CR Subject Classification (1998): I.2.11, I.2, C.2.4, H.5.2-3

ISSN 0302-9743
ISBN 3-540-24526-X Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11384885 06/3142 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence 3366

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Lecture Notes in Artificial Intelligence (LNAI)

Vol. 3403: B. Ganter, R. Godin (Eds.), Formal Concept Analysis. XI, 419 pages. 2005.

Vol. 3398: D.-K. Baik (Ed.), Systems Modeling and Simulation: Theory and Applications. XIV, 733 pages. 2005.

Vol. 3397: T.G. Kim (Ed.), Artificial Intelligence and Simulation. XV, 711 pages. 2005.

Vol. 3366: I. Rahwan, P. Moraitis, C. Reed (Eds.), Argumentation in Multi-Agent Systems. XII, 263 pages. 2005.

Vol. 3359: G. Grieser, Y. Tanaka (Eds.), Intuitive Human Interfaces for Organizing and Accessing Intellectual Assets. XIV, 257 pages. 2005.

Vol. 3346: R.H. Bordini, M. Dastani, J. Dix, A.E.F. Seghrouchni (Eds.), Programming Multi-Agent Systems. XIV, 249 pages. 2005.

Vol. 3345: Y. Cai (Ed.), Ambient Intelligence for Scientific Discovery. XII, 311 pages. 2005.

Vol. 3339: G.I. Webb, X. Yu (Eds.), AI 2004: Advances in Artificial Intelligence. XXII, 1272 pages. 2004.

Vol. 3336: D. Karagiannis, U. Reimer (Eds.), Practical Aspects of Knowledge Management. X, 523 pages. 2004.

Vol. 3327: Y. Shi, W. Xu, Z. Chen (Eds.), Data Mining and Knowledge Management. XIII, 263 pages. 2004.

Vol. 3315: C. Lemaître, C.A. Reyes, J.A. González (Eds.), Advances in Artificial Intelligence – IBERAMIA 2004. XX, 987 pages. 2004.

Vol. 3303: J.A. López, E. Benfenati, W. Dubitzky (Eds.), Knowledge Exploration in Life Science Informatics. X, 249 pages. 2004.

Vol. 3275: P. Perner (Ed.), Advances in Data Mining. VIII, 173 pages. 2004.

Vol. 3265: R.E. Frederking, K.B. Taylor (Eds.), Machine Translation: From Real Users to Research. XI, 392 pages. 2004.

Vol. 3264: G. Paliouras, Y. Sakakibara (Eds.), Grammatical Inference: Algorithms and Applications. XI, 291 pages. 2004.

Vol. 3259: J. Dix, J. Leite (Eds.), Computational Logic in Multi-Agent Systems. XII, 251 pages. 2004.

Vol. 3257: E. Motta, N.R. Shadbolt, A. Stutt, N. Gibbins (Eds.), Engineering Knowledge in the Age of the Semantic Web. XVII, 517 pages. 2004.

Vol. 3249: B. Buchberger, J.A. Campbell (Eds.), Artificial Intelligence and Symbolic Computation. X, 285 pages. 2004.

Vol. 3248: K.-Y. Su, J.-i. Tsujii, J.-H. Lee, O.Y. Kwong (Eds.), Natural Language Processing – IJCNLP 2004. XVIII, 817 pages. 2005.

Vol. 3245: E. Suzuki, S. Arikawa (Eds.), Discovery Science. XIV, 430 pages. 2004.

Vol. 3244: S. Ben-David, J. Case, A. Maruoka (Eds.), Algorithmic Learning Theory. XIV, 505 pages. 2004.

Vol. 3238: S. Biundo, T. Frühwirth, G. Palm (Eds.), KI 2004: Advances in Artificial Intelligence. XI, 467 pages. 2004.

Vol. 3230: J.L. Vicedo, P. Martínez-Barco, R. Muñoz, M. Saiz Noeda (Eds.), Advances in Natural Language Processing. XII, 488 pages. 2004.

Vol. 3229: J.J. Alferes, J. Leite (Eds.), Logics in Artificial Intelligence. XIV, 744 pages. 2004.

Vol. 3228: M.G. Hinchey, J.L. Rash, W.F. Truszkowski, C.A. Rouff (Eds.), Formal Approaches to Agent-Based Systems. VIII, 290 pages. 2004.

Vol. 3215: M.G.. Negoita, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part III. LVII, 906 pages. 2004.

Vol. 3214: M.G.. Negoita, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part II. LVIII, 1302 pages. 2004.

Vol. 3213: M.G.. Negoita, R.J. Howlett, L.C. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part I. LVIII, 1280 pages. 2004.

Vol. 3209: B. Berendt, A. Hotho, D. Mladenic, M. van Someren, M. Spiliopoulou, G. Stumme (Eds.), Web Mining: From Web to Semantic Web. IX, 201 pages. 2004.

Vol. 3206: P. Sojka, I. Kopecek, K. Pala (Eds.), Text, Speech and Dialogue. XIII, 667 pages. 2004.

Vol. 3202: J.-F. Boulicaut, F. Esposito, F. Giannotti, D. Pedreschi (Eds.), Knowledge Discovery in Databases: PKDD 2004. XIX, 560 pages. 2004.

Vol. 3201: J.-F. Boulicaut, F. Esposito, F. Giannotti, D. Pedreschi (Eds.), Machine Learning: ECML 2004. XVIII, 580 pages. 2004.

Vol. 3194: R. Camacho, R. King, A. Srinivasan (Eds.), Inductive Logic Programming. XI, 361 pages. 2004.

Vol. 3192: C. Bussler, D. Fensel (Eds.), Artificial Intelligence: Methodology, Systems, and Applications. XIII, 522 pages. 2004.

Vol. 3191: M. Klusch, S. Ossowski, V. Kashyap, R. Unland (Eds.), Cooperative Information Agents VIII. XI, 303 pages. 2004.

Vol. 3187: G. Lindemann, J. Denzinger, I.J. Timm, R. Unland (Eds.), Multiagent System Technologies. XIII, 341 pages. 2004.

Vol. 3176: O. Bousquet, U. von Luxburg, G. Rätsch (Eds.), Advanced Lectures on Machine Learning. IX, 241 pages. 2004.

Vol. 3171: A.L.C. Bazzan, S. Labidi (Eds.), Advances in Artificial Intelligence – SBIA 2004. XVII, 548 pages. 2004.

Vol. 3159: U. Visser, Intelligent Information Integration for the Semantic Web. XIV, 150 pages. 2004.

Vol. 3157: C. Zhang, H. W. Guesgen, W.K. Yeap (Eds.), PRICAI 2004: Trends in Artificial Intelligence. XX, 1023 pages. 2004.

Vol. 3155: P. Funk, P.A. González Calero (Eds.), Advances in Case-Based Reasoning. XIII, 822 pages. 2004.

Vol. 3139: F. Iida, R. Pfeifer, L. Steels, Y. Kuniyoshi (Eds.), Embodied Artificial Intelligence. IX, 331 pages. 2004.

Vol. 3131: V. Torra, Y. Narukawa (Eds.), Modeling Decisions for Artificial Intelligence. XI, 327 pages. 2004.

Vol. 3127: K.E. Wolff, H.D. Pfeiffer, H.S. Delugach (Eds.), Conceptual Structures at Work. XI, 403 pages. 2004.

Vol. 3123: A. Belz, R. Evans, P. Piwek (Eds.), Natural Language Generation. X, 219 pages. 2004.

Vol. 3120: J. Shawe-Taylor, Y. Singer (Eds.), Learning Theory. X, 648 pages. 2004.

Vol. 3097: D. Basin, M. Rusinowitch (Eds.), Automated Reasoning. XII, 493 pages. 2004.

Vol. 3071: A. Omicini, P. Petta, J. Pitt (Eds.), Engineering Societies in the Agents World. XIII, 409 pages. 2004.

Vol. 3070: L. Rutkowski, J. Siekmann, R. Tadeusiewicz, L.A. Zadeh (Eds.), Artificial Intelligence and Soft Computing - ICAISC 2004. XXV, 1208 pages. 2004.

Vol. 3068: E. André, L. Dybkjær, W. Minker, P. Heisterkamp (Eds.), Affective Dialogue Systems. XII, 324 pages. 2004.

Vol. 3067: M. Dastani, J. Dix, A. El Fallah-Seghrouchni (Eds.), Programming Multi-Agent Systems. X, 221 pages. 2004.

Vol. 3066: S. Tsumoto, R. Słowiński, J. Komorowski, J.W. Grzymała-Busse (Eds.), Rough Sets and Current Trends in Computing. XX, 853 pages. 2004.

Vol. 3065: A. Lomuscio, D. Nute (Eds.), Deontic Logic in Computer Science. X, 275 pages. 2004.

Vol. 3060: A.Y. Tawfik, S.D. Goodwin (Eds.), Advances in Artificial Intelligence. XIII, 582 pages. 2004.

Vol. 3056: H. Dai, R. Srikant, C. Zhang (Eds.), Advances in Knowledge Discovery and Data Mining. XIX, 713 pages. 2004.

Vol. 3055: H. Christiansen, M.-S. Hacid, T. Andreasen, H.L. Larsen (Eds.), Flexible Query Answering Systems. X, 500 pages. 2004.

Vol. 3048: P. Faratin, D.C. Parkes, J.A. Rodríguez-Aguilar, W.E. Walsh (Eds.), Agent-Mediated Electronic Commerce V. XI, 155 pages. 2004.

Vol. 3040: R. Conejo, M. Urretavizcaya, J.-L. Pérez-de-la-Cruz (Eds.), Current Topics in Artificial Intelligence. XIV, 689 pages. 2004.

Vol. 3035: M.A. Wimmer (Ed.), Knowledge Management in Electronic Government. XII, 326 pages. 2004.

Vol. 3034: J. Favela, E. Menasalvas, E. Chávez (Eds.), Advances in Web Intelligence. XIII, 227 pages. 2004.

Vol. 3030: P. Giorgini, B. Henderson-Sellers, M. Winikoff (Eds.), Agent-Oriented Information Systems. XIV, 207 pages. 2004.

Vol. 3029: B. Orchard, C. Yang, M. Ali (Eds.), Innovations in Applied Artificial Intelligence. XXI, 1272 pages. 2004.

Vol. 3025: G.A. Vouros, T. Panayiotopoulos (Eds.), Methods and Applications of Artificial Intelligence. XV, 546 pages. 2004.

Vol. 3020: D. Polani, B. Browning, A. Bonarini, K. Yoshida (Eds.), RoboCup 2003: Robot Soccer World Cup VII. XVI, 767 pages. 2004.

Vol. 3012: K. Kurumatani, S.-H. Chen, A. Ohuchi (Eds.), Multi-Agents for Mass User Support. X, 217 pages. 2004.

Vol. 3010: K.R. Apt, F. Fages, F. Rossi, P. Szeredi, J. Váncza (Eds.), Recent Advances in Constraints. VIII, 285 pages. 2004.

Vol. 2990: J. Leite, A. Omicini, L. Sterling, P. Torroni (Eds.), Declarative Agent Languages and Technologies. XII, 281 pages. 2004.

Vol. 2980: A. Blackwell, K. Marriott, A. Shimojima (Eds.), Diagrammatic Representation and Inference. XV, 448 pages. 2004.

Vol. 2977: G. Di Marzo Serugendo, A. Karageorgos, O.F. Rana, F. Zambonelli (Eds.), Engineering Self-Organising Systems. X, 299 pages. 2004.

Vol. 2972: R. Monroy, G. Arroyo-Figueroa, L.E. Sucar, H. Sossa (Eds.), MICAI 2004: Advances in Artificial Intelligence. XVII, 923 pages. 2004.

Vol. 2969: M. Nickles, M. Rovatsos, G. Weiss (Eds.), Agents and Computational Autonomy. X, 275 pages. 2004.

Vol. 2961: P. Eklund (Ed.), Concept Lattices. IX, 411 pages. 2004.

Vol. 2953: K. Konrad, Model Generation for Natural Language Interpretation and Analysis. XIII, 166 pages. 2004.

Vol. 2934: G. Lindemann, D. Moldt, M. Paolucci (Eds.), Regulated Agent-Based Social Systems. X, 301 pages. 2004.

Vol. 2930: F. Winkler (Ed.), Automated Deduction in Geometry. VII, 231 pages. 2004.

Vol. 2926: L. van Elst, V. Dignum, A. Abecker (Eds.), Agent-Mediated Knowledge Management. XI, 428 pages. 2004.

Vol. 2923: V. Lifschitz, I. Niemelä (Eds.), Logic Programming and Nonmonotonic Reasoning. IX, 365 pages. 2003.

Vol. 2915: A. Camurri, G. Volpe (Eds.), Gesture-Based Communication in Human-Computer Interaction. XIII, 558 pages. 2004.

Vol. 2913: T.M. Pinkston, V.K. Prasanna (Eds.), High Performance Computing - HiPC 2003. XX, 512 pages. 2003.

Vol. 2903: T.D. Gedeon, L.C.C. Fung (Eds.), AI 2003: Advances in Artificial Intelligence. XVI, 1075 pages. 2003.

Vol. 2902: F.M. Pires, S.P. Abreu (Eds.), Progress in Artificial Intelligence. XV, 504 pages. 2003.

Vol. 2892: F. Dau, The Logic System of Concept Graphs with Negation. XI, 213 pages. 2003.

Vol. 2891: J. Lee, M. Barley (Eds.), Intelligent Agents and Multi-Agent Systems. X, 215 pages. 2003.

Vol. 2882: D. Veit, Matchmaking in Electronic Markets. XV, 180 pages. 2003.

Vol. 2872: G. Moro, C. Sartori, M.P. Singh (Eds.), Agents and Peer-to-Peer Computing. XII, 205 pages. 2004.

Preface

The theory of argumentation is a rich, interdisciplinary area of research lying across philosophy, communication studies, linguistics, and psychology (at least). Its techniques and results have found a wide range of applications in both theoretical and practical branches of artificial intelligence and computer science. Several theories of argumentation with various semantics have been proposed in the literature. Multi-agent systems theory has picked up argument-inspired approaches and specifically argumentation-theoretic results from many different areas.

The community of researchers in argumentation and multi-agent systems is currently presented with a unique opportunity to integrate the various understandings of argument into a coherent and core part of the functioning of autonomous computational systems. The benefits range from extended semantics of arguments construed as relationships between epistemic atoms, through conversation protocols for argumentation with serendipitous information exchange, to models of dialectical practical reasoning, both intra- and inter-agent (and a mixture of the two). In all these cases argumentation is used to structure knowledge representation, reasoning and agent interaction, and offers a potential means of better integrating these disparate problems.

In recognition of this increasing interest, the 1st International Workshop on Argumentation in Multi-agent Systems (ArgMAS) was conceived. The workshop was the first forum that brought together researchers interested in applying argumentation to problems faced by the Autonomous Agents and Multi-agent Systems (AAMAS) community. Hence, the workshop was held in conjunction with the 3rd International AAMAS Conference, in July 2004 at Columbia University, New York. The workshop received 20 full-paper submissions and 2 position statements, which was a very encouraging sign for a new workshop. After a thorough reviewing process by at least 2 anonymous referees per paper, 13 full papers were selected for presentation at the workshop. The workshop also included an invited talk by Prof. Jonathan Adler from the Faculty of Philosophy, City University of New York. In this volume, we included revised and expanded versions of the 13 workshop papers. In addition, we included 4 invited contributions, which range from relevant papers that appeared at the main AAMAS conference to contributions from prominent researchers in the field who did not make it to the workshop. Invited contributions were also fully refereed, either by the AAMAS or ArgMAS reviewers. As a result, the book provides a strong representation of the state of the art in the emerging field. Papers range from specific technical contributions to discussions of overarching issues in the area.

The papers were roughly divided into the following main themes:

- Foundations of dialogues
- Belief revision
- Persuasion and deliberation

- Negotiation
- Strategic issues

Although these topics are not completely distinct, they indicate some main directions of research. We have therefore arranged the papers in the book according to these themes.

The first five papers (Part I) address foundational issues in argumentation-based multi-agent dialogues. The first paper (by *Simon Parsons*, *Peter McBurney* and *Michael Wooldridge*) sets down some preliminary but important steps towards a meta-theory of inter-agent dialogues by examining different classes of protocols and how they may lead to different interaction outcomes. The next paper (by *Chris Reed* and *Doug Walton*) looks at formalizing and implementing argumentation schemes, a form of non-deductive reasoning. This is followed by another paper (by *Simon Wells* and *Chris Reed*) which explores the specification of formal dialectic Hamblin-type systems, and presents an implemented system that makes use of the formal framework. The fourth paper (by *Jamal Bentahar*, *Bernard Moulin*, *John-Jules Ch. Meyer* and *Brahim Chaib-draa*) provides an approach based on modal logic for providing semantics for commitments during argumentation dialogues. This paper was invited after being accepted for presentation at the main conference. The last paper in Part I (by *Antonis Kakas*, *Nicolas Maudet* and *Pavlos Moraitis*) explores the interplay between dialogue protocols and agent internal strategies, and analyzes these within a single theoretical framework.

Part II focuses on the use of argumentation as a reasoning mechanism for revising beliefs in the context of a changing environment. The first paper in this section (by *Fabio Paglieri* and *Cristiano Castelfranchi*) provides the reader with a good scoping of the research field of the workshop. In particular, it argues that belief revision and argumentation are complementary components of belief change in multi-agent systems. Next, a specific model for argumentation-based belief revision is presented in a separate paper by *Marcela Capobianco*, *Carlos I. Chesñevar* and *Guillermo R. Simari*. The final paper in this section is an invited contribution (by *Gerard Vreeswijk*) on the relationship between argumentation-based reasoning and Bayesian probabilistic inference. This contribution promises to open up new avenues of research to bridge the gap between the symbolic and probabilistic views of communication.

Part III of this volume presents three contributions to multi-agent persuasion and deliberation dialogues. The first paper (by *Jamal Bentahar*, *Bernard Moulin* and *Brahim Chaib-draa*) presents a persuasion dialogue game protocol and studies the dynamics of the commitments of agents using the protocol. The following two papers contribute to deliberation dialogues, interactions where participants jointly decide on a course of action. The first of those (by *Katie Atkinson*, *Trevor Bench-Capon* and *Peter McBurney*) presents a dialogue game protocol for deliberation dialogues. This is followed by another paper (by *Peter McBurney* and *Simon Parsons*) which proposes a denotational semantics for deliberation dialogues, based on mathematical category theory.

Part IV concentrates on argumentation-based negotiation dialogues, an area receiving increasing interest in the multi-agent systems community. The first paper (by *Iyad Rahwan*, *Liz Sonenberg* and *Peter McBurney*) discusses the difference between argumentation-based negotiation and traditional bargaining, in which agents simply exchange offers. This is followed by a paper by *Leila Amgoud* and *Souhila Kaci*, who present an argumentation-based approach to generate desires and goals. This approach has potential benefit for negotiation dialogues as it provides a means for allowing agents to influence each others' preferences during negotiation. The third paper in this part (by *Sabyasachi Saha* and *Sandip Sen*) presents an approach for argumentation-based negotiation based on Bayesian networks. This is a slightly different treatment from that presented in the paper by *Gerard Vreeswijk* in Part II, since it uses Bayesian networks in order to model the negotiation opponent's behavior. The last paper, by *Fernando A. Tohmé* and *Guillermo R. Simari*, presents a framework for negotiation based on defeasible logic programming (DeLP) augmented with utility functions.

Finally, Part V contains papers that explore various issues related to agent decision-making in dialogues, i.e., their strategies. The first paper (by *Nishan C. Karunatillake* and *Nicholas R. Jennings*) uses empirical simulation to investigate whether and when argumentation improves negotiation. They demonstrate that argumentation is useful when resources are relatively scarce, but provide marginal benefit when resources are abundant. The second paper (by *Elizabeth Sklar*, *Simon Parsons* and *Mathew Davies*) explores the issue of lying in multi-agent dialogues and shows that lying can be useful, and even acceptable, in certain circumstances.

Together the papers in the five parts capture the current landscape of uses of argumentation in multi-agent systems. As a young and dynamic field of research, fresh with vitality, advances are being made extremely rapidly, but nevertheless there are some few trends that are worth identifying in trying to understand where the research is heading. Perhaps the first and most striking is that there is an increasing appeal from more informal areas of argumentation theory. Thus rhetoric, with its focus on audiences, values and context-dependence, is becoming more visible as agents become more sophisticated in their communication structures and reasoning capabilities. The more complex such capabilities become, the more susceptible those systems become to rhetorical techniques. Similarly, argumentation schemes, which encompass a wide range of humanistic reasoning techniques, are being harnessed for internal agent reasoning and inter-agent communication. As the structure of agent knowledge bases becomes more refined, the reasoning techniques that can be leveraged become more detailed and more specific.

Another clear trend is the emergence of the need for objective comparisons between systems. In some cases, such evaluation can be carried out using tools from earlier multi-agent systems research or distributed computing. Yet, much more commonly, the tools for evaluation simply do not exist and need building from scratch. As the range of argumentation-based techniques for reasoning and

communicating expands, benchmarking and evaluation will become an increasingly important requirement in comparing and assessing those techniques.

A very important research trend, which we are only beginning to see glimpses of, is the integration of argumentation-theoretic and economic-theoretic conceptions of rationality. Attempts to integrate notions of economic preference (e.g., via the notion of *utility*) into argumentation systems is an important step towards integration.

Finally, and looking to the longer term, we foresee the emergence of richer argumentation models such as those that move away from the so-called "standard treatment" (such as formalizations of Toulmin's model). These will be driven by the limitations of expressivity identified in dialectical models (e.g., refutations versus negations; distinctions between undercutting and rebutting; and distinctions between warrants and implications). As agent reasoning becomes more sophisticated, the limits of the propositional model come ever more to the fore. Perhaps it is the ArgMAS community that will be at the vanguard of engineering solutions that tackle induction, categorical syllogism, the interrogative and imperative, and a whole host of Aristotelian basic concepts that might yield concrete computational gains in implemented agent systems.

We conclude this preface by extending our gratitude to the members of the steering committee, members of the program committee, and the auxiliary reviewers, who together helped make the ArgMAS workshop a success. We also thank the authors for their enthusiasm in submitting papers to the workshop, and for revising their papers on time for inclusion in this book.

October 2004

Iyad Rahwan, Pavlos Moraitis, and Chris Reed
Program Chairs
ArgMAS 2004

Organization

Program Chairs

Iyad Rahwan	University of Melbourne, Australia The British University in Dubai, UAE
Pavlos Moraitis	University of Cyprus, Cyprus
Chris Reed	University of Dundee, UK

ArgMAS Steering Committee

Antonis Kakas	University of Cyprus, Cyprus
Nicolas Maudet	Université Paris Dauphine, France
Peter McBurney	University of Liverpool, UK
Pavlos Moraitis	University of Cyprus, Cyprus
Simon Parsons	City University of New York, USA
Iyad Rahwan	University of Melbourne, Australia
Chris Reed	University of Dundee, UK

Program Committee

Leila Amgoud	IRIT, France
Frank Dignum	Utrecht University, Netherlands
Rogier van Eijk	Utrecht University, Netherlands
Antonis Kakas	University of Cyprus, Cyprus
Nicolas Maudet	Université Paris Dauphine, France
Peter McBurney	University of Liverpool, UK
Pavlos Moraitis	University of Cyprus, Cyprus
Xavier Parent	King's College, London, UK
Simon Parsons	City University of New York, USA
Henry Prakken	Utrecht University, Netherlands
Iyad Rahwan	University of Melbourne, Australia
Chris Reed	University of Dundee, UK
Carles Sierra	IIIA, Spain
Paolo Torroni	Università di Bologna, Italy
Bart Verheij	Maastricht University, Netherlands
Gerard Vreeswijk	Utrecht University, Netherlands
Mike Wooldridge	University of Liverpool, UK

Auxiliary Referees

Evelina Lamma
Paola Mello

Table of Contents

Part I: Foundations of Dialogues

Part II: Belief Revision

Part III: Persuasion & Deliberation

Part IV: Negotiation

Part V: Strategic Issues

Some Preliminary Steps Towards a Meta-theory for Formal Inter-agent Dialogues

Simon Parsons[1], Peter McBurney[2], and Michael Wooldridge[2]

[1] Department of Computer and Information Science, Brooklyn College, City University of New York, 2900 Bedford Avenue, Brooklyn, New York, NY 11210, USA
parsons@sci.brooklyn.cuny.edu
[2] Department of Computer Science, University of Liverpool, Chadwick Building, Liverpool L69 7ZF, UK
{p.j.mcburney, m.j.wooldridge}@csc.liv.ac.uk

Abstract. This paper investigates the properties of argumentation-based dialogues between agents. It takes a previously defined system by which agents can trade arguments, and examines how different classes of protocols for this kind of interaction can have profoundly different outcomes. Studying such classes of protocol, rather than individual protocols as has been done previously, allows us to start to develop a *meta-theory* of this class of interactions.

1 Introduction

Research into the theoretical properties of protocols for multi-agent interaction can be crudely divided into two camps. The first camp is broadly characterised by the application of game and economic theory to understanding the properties of multi-agent protocols; this camp includes, for example, research on auction protocols and algorithmic mechanism design [12]. The second camp may be broadly characterised by an understanding of agents as practical reasoning systems, which interact in order to to resolve differences of opinion and conflicts of interest; to work together to resolve dilemmas or find proofs; or simply to inform each other of pertinent facts. As work in the former camp has been informed by game and economic theory, so work in this latter camp has been informed in particular by research in the area of *argumentation* and *dialogue games*. Examples of argumentation-based approaches to multi-agent dialogues include the work of Dignum *et al.* [4], Kraus [13], Reed [20], Schroeder *et al.* [21] and Sycara [22].

The work of Walton and Krabbe has been particularly influential in argumentation-based dialogue research [23]. They developed a typology for inter-personal dialogue which identifies six primary types of dialogues and three mixed types. The categorization is based upon: what information the participants each have at the commencement of the dialogue (with regard to the topic of discussion); what goals the individual participants have; and what goals are shared by the participants, goals we may view as those of the dialogue itself. This *dialogue game* view of dialogues overlaps with work on conversation policies (see, for example, [3, 6]), but differs in considering the entire dialogue rather than dialogue segments. As defined by Walton and Krabbe, the three types of dialogue

I. Rahwan et al. (Eds.): ArgMAS 2004, LNAI 3366, pp. 1–18, 2005.

we have considered in our previous work are: *Information-Seeking Dialogues* (where one participant seeks the answer to some question(s) from another participant, who is believed by the first to know the answer(s)); *Inquiry Dialogues* (where the participants collaborate to answer some question or questions whose answers are not known to any one participant); and *Persuasion Dialogues* (where one party seeks to persuade another party to adopt a belief or point-of-view he or she does not currently hold). Persuasion dialogues begin with one party supporting a particular statement which the other party to the dialogue does not, and the first seeks to convince the second to adopt the proposition. The second party may not share this objective.

Our previous work investigated capturing these types of dialogue using a formal model of argumentation [2], and the basic properties and complexity of such dialogues [16]. Most recently, we have looked at how the outcomes of these dialogues can depend upon the order in which agents make utterances [17]. Here we extend this investigation, by moving from the study of particular protocols to the study of *classes of protocols*, and the properties of those classes. These results, then, are (very preliminary) results about the *meta-theory* of argumentation-based dialogues. The advantage of this change in perspective is that our results are robust—they hold for a wider range of possible dialogues—and more wide-reaching that we have been able to obtain hitherto, permitting a more complete analysis of argumentation-based dialogues. Note that, despite the fact that the types of dialogue we are considering are drawn from the analysis of human dialogues, we are only concerned here with dialogues between *artificial* agents. Unlike Grosz and Sidner [10] for example, we choose to focus in this way in order to simplify our task—dealing with artificial languages avoids much of the complexity of natural language dialogues.

2 Background

In this section, we briefly introduce the formal system of argumentation that underpins our approach [1], a system that extends Dung's [5] with preferences. We start with a (possibly inconsistent) knowledge base Σ with no deductive closure. We assume Σ contains formulas of a propositional language $\mathcal{L}$, that $\vdash$ stands is the classical inference relation, and $\equiv$ stands for logical equivalence. An argument is a proposition and the set of formulae from which it can be inferred:

Definition 1. *An* argument *is a pair* $A = (H, h)$ *where* h *is a formula of* $\mathcal{L}$ *and* H *a subset of* Σ *such that:*

1. H *is consistent;*
2. $H \vdash h$*; and*
3. H *is minimal, so no proper subset of* H *satisfying both (1) and (2) exists.*

H *is called the* support *of* A*, written* $H = Support(A)$ *and* h *is the* conclusion *of* A*, written* $h = Conclusion(A)$*.*

We thus talk of h being *supported* by the argument (H, h)

In general, since Σ is inconsistent, arguments in $\mathcal{A}(\Sigma)$, the set of all arguments which can be made from Σ, will conflict, and we make this idea precise with the notion of *undercutting*:

Definition 2. *Let A_1 and A_2 be two arguments of $\mathcal{A}(\Sigma)$. A_1* undercuts *A_2 iff $\exists h \in Support(A_2)$ such that $h \equiv \neg Conclusion(A_1)$.*

In other words, an argument is undercut iff there is another argument which has as its conclusion the negation of an element of the support for the first argument.

To capture the fact that some facts are more strongly believed than others, we assume that any set of facts has a preference order over it. We suppose that this ordering derives from the fact that the knowledge base Σ is stratified into non-overlapping sets $\Sigma_1, \ldots, \Sigma_n$ such that facts in Σ_i are all equally preferred, and are more preferred than those in Σ_j where $j > i$. The preference level of a nonempty subset H of Σ, $level(H)$, is the number of the highest numbered layer which has a member in H.

Definition 3. *Let A_1 and A_2 be two arguments in $\mathcal{A}(\Sigma)$. A_1 is* preferred *to A_2 according to $Pref$, $Pref(A_1, A_2)$, iff $level(Support(A_1)) \leq level(Support(A_2))$.*

By $\gg^{Pref}$, we denote the strict pre-order associated with $Pref$. If A_1 is preferred to A_2, we say that A_1 is *stronger* than A_2[1]. We can now define the argumentation system we will use:

Definition 4. *An* argumentation system *(AS) is a triple $\langle \mathcal{A}(\Sigma), Undercut, Pref \rangle$ such that:*

- *$\mathcal{A}(\Sigma)$ is a set of the arguments built from Σ,*
- *$Undercut$ is a binary relation representing the defeat relationship between arguments, $Undercut \subseteq \mathcal{A}(\Sigma) \times \mathcal{A}(\Sigma)$, and*
- *$Pref$ is a (partial or complete) preordering on $\mathcal{A}(\Sigma) \times \mathcal{A}(\Sigma)$.*

The preference order makes it possible to distinguish different types of relation between arguments:

Definition 5. *Let A_1, A_2 be two arguments of $\mathcal{A}(\Sigma)$.*

- *If A_2 undercuts A_1 then A_1* defends itself *against A_2 iff $A_1 \gg^{Pref} A_2$. Otherwise, A_1* does not defend itself.
- *A set of arguments $\mathcal{S}$* defends *A iff: $\forall$ B undercuts A and A does not defend itself against B then $\exists$ $C \in \mathcal{S}$ such that C undercuts B and B does not defend itself against C.*

We write $C_{Undercut, Pref}$ to denote the set of all non-undercut arguments and arguments defending themselves against all their undercutting arguments. The set $\underline{\mathcal{S}}$ of acceptable arguments of the argumentation system $\langle \mathcal{A}(\Sigma), Undercut, Pref \rangle$ is the least fixpoint of a function $\mathcal{F}$ [1]:

$$\mathcal{S} \subseteq \mathcal{A}(\Sigma)$$
$$\mathcal{F}(\mathcal{S}) = \{(H, h) \in \mathcal{A}(\Sigma) \mid (H, h) \text{ is defended by } \mathcal{S}\}$$

[1] We acknowledge that this model of preferences is rather restrictive and in the future intend to work to relax it.